ZooKeys 428: 97–108 (2014)
doi: 10.3897/zookeys.428.7366
www.zookeys.org

SHORT COMMUNICATION

A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

A set of multi-entry identification keys to African frugivorous flies (Diptera, Tephritidae)

Massimiliano Virgilio[1], Ian White[2], Marc De Meyer[1]

1 *Royal Museum for Central Africa, Leuvensesteenweg 13, B3080 Tervuren, Belgium* **2** *Buxton, UK; previously The Natural History Museum, Cromwell Road, London, SW7 5BD*

Corresponding author: *Massimiliano Virgilio* (massimiliano.virgilio@africamuseum.be)

Academic editor: *Rudolf Meier* | Received 22 February 2014 | Accepted 11 July 2014 | Published 24 July 2014

http://zoobank.org/0C5847D5-EC3D-4CE6-9579-282846092D99

Citation: Virgilio M, White I, De Meyer M (2014) A set of multi-entry identification keys to African frugivorous flies (Diptera, Tephritidae). ZooKeys 428: 97–108. doi: 10.3897/zookeys.428.7366

Keywords

Tephritidae, Africa, identification key

Introduction

Tephritid fruit flies, or "true" fruit flies (Diptera, Tephritidae) include approximately 500 genera and 4800 valid species (Norrbom 2004), whose vast majority (95%) is represented by phytophagous species (reviewed in Aluja and Norrbom 1999). Among them, frugivorous flies represent approximately 25–30% of all tephritid species, occur in tropical and temperate regions of all continents except the Antarctic and are predominantly distributed in five main genera (*Anastrepha* Schiner, *Rhagoletis* Loew, *Ceratitis* MacLeay, *Dacus* Fabricius and *Bactrocera* Macquart). Frugivorous tephritids attack healthy fruit still on the tree. The larvae develop inside the fruit, feed on the plant tissues, and complete their developmental cycle in the soil. A relatively limited number (approximately 100) of frugivorous species are phytophagous pests whose larvae attack pulp and/or seeds of cultivated fruits and crops of agricultural importance. In Africa, damage on commercial fruits and crops is caused mainly by polyphagous species belonging to the genera *Ceratitis*, *Dacus* and *Bactrocera* (De Meyer et al. 2008;

Copyright *Massimiliano Virgilio et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

White 2006). Other African and closely related genera with fewer taxa are *Capparimyia* Bezzi, *Carpophthoromyia* Austen, *Neoceratitis* Hendel, *Perilampsis* Bezzi and *Trirhithrum* Bezzi, which also include some species of economic significance.

Currently, identification of tephritid flies is a specialized task largely performed by a restricted pool of experienced taxonomists, a group that is constantly becoming smaller due to the well-known problems related to the general loss of taxonomical expertise on insects as well as on most taxonomic groups (Carvalho et al. 2007; Wilson 2000). In the last few decades, globalisation of fruit trade and transport (Aluja and Mangan 2008; Malacrida et al. 2007) has made the need for swift, reliable and accurate identification methods for frugivorous flies even more urgent. For example, in 1995, the erroneous identification of *B. zonata* as *B. pallidus* in Egypt produced a three-year delay on implementation of phytosanitary measures and resulted in serious damage to the agricultural productivity of the whole Alexandria region.

The morphological identification of African tephritids largely depends on the use of classical single-entry (dichotomous) keys. These keys are available for most African genera (e.g., White 2006), with the important exception of the genus *Ceratitis*, whose species can only be identified through separate subgeneric keys (De Meyer 1996, 1998, 2000; De Meyer and Freidberg 2006). The main disadvantage of single-entry keys is that species identification inevitably fails whenever the user is not able to select any of the dichotomous character states listed in the key (e.g., due to his inadequate taxonomic expertise, lack of clarity of the key, damaged specimen, etc.). Additionally, the specific terminology used in published keys represents a serious obstacle for non-specialist users who are not particularly acquainted with insect morphology and taxonomy. For these reasons, obtaining the taxonomical expertise that is necessary to identify tephritids using the above mentioned tools has never been an easy task, particularly for African scientists who can only rely on a limited number of comprehensive reference collections in the continent (as, for example, the South African National Collection of Insects, Pretoria - South Africa, the National Museums of Kenya, Nairobi - Kenya, or the International Institute of Tropical Agriculture (IITA), Cotonou - Benin). Molecular techniques represent a partial solution to counteract loss of taxonomical expertise on tephritid flies. DNA barcoding has been proposed as a relatively rapid and effective tool for the identification of fruit flies (Armstrong and Ball 2005). Yet, despite the availability of relatively large reference libraries of DNA barcodes for tephritid fruit flies, this method is still not routinely used for identification mainly due to shortcomings such as the difficulty of resolving important species complexes (Smit et al. 2013; Virgilio et al. 2010) and the incompleteness of reference libraries (Virgilio et al. 2012).

To try and reduce the effects of some of the aforementioned issues, we developed a set of freely available multi-entry identification keys for African fruit flies. The keys provide a professional identification tool that is also accessible to non-specialised morphologists (*i.e.*, people that might be interested in fruit fly identification such as students, technicians, agronomists, quarantine officers, ecologists, farmers, molecular biologists, etc.). Matrixes containing scores for 340 characters from 400 African spe-

cies belonging to the genera *Bactrocera*, *Capparimyia*, *Carpophthoromyia*, *Ceratitis*, *Dacus*, *Neoceratitis*, *Perilampsis* and *Trirhithrum* were compiled from data sets that were used within the framework of previous taxonomic revisions (De Meyer 1996, 1998, 2000, 2006, 2009; De Meyer and Freidberg 2005, 2006, 2012; White et al. 2003; White 2006; White and Goodger 2009). Scores were transferred into seven separate data sets, imported into LUCID 3.5 (www.lucidcentral.org) and used as the main data sources for the multi-entry identification keys. Species lists and morphological characters were then revised and optimised in order to include only (a) species with valid names under the International Code of Zoological Nomenclature and (b) characters including at least two character states in congeneric species. This generated 7 matrixes with a total of 68352 entries. Additionally, a "pre-key" for genus designation was built *ex novo* by selecting a set of 23 characters that were deemed to be informative for generic separation. A total of 390 taxa were included in seven identification keys for species identification within genus or genus group (*Bactrocera* + *Dacus*, *Capparimyia*, *Carpophthoromyia*, *Ceratitis*, *Neoceratitis*, *Perilampsis*, *Trirhithrum*). For each genus, species of economic importance were assigned to a separate subset (see below).

Different character sets were considered for each genus (range 11-90 characters and 22–204 character states). The complete lists of species, characters, character states and dependencies considered for each key are provided as supplementary files (SF1, SF2). Each character state was scored in LUCID as either "present and common" or "absent" (other options such as "present but rare", "common and misinterpreted" etc. were not implemented). The "not scoped" option was used to generate unfolding keys, *i.e.* keys with characters that are initially not shown but appear only when a pre-defined subset of species remains to be identified. We built unfolding keys whenever character scores were only available for subsets of a maximum of 5 congeneric taxa. Dependencies between characters were also generated. Positive dependencies were defined whenever a character was only meaningful in relation to a previously defined character state (*e.g.* in the *Ceratitis* key, the character "number of frontal setae" is positively dependent from the character state "frontal setae: yes"). Conversely, negative dependencies were generated to discard characters that were not meaningful after a previous character state was selected (*e.g.* in the *Ceratitis* key, the character "females, aculeus tip with small notch" is negatively dependent on the character state "sex: male"). To facilitate identification, characters were grouped into head, thorax, wings, legs and abdomen character sets. The character "sex" was always placed first, in order to reduce the character list by discarding all negative dependencies controlled by the character states "male" and "female".

We considered that the number of morphological characters used in the largest identification keys (*i.e.* keys to *Bactrocera*/*Dacus*, *Ceratitis*, *Trirhithrum*) might also represent an obstacle to non-specialists. Hence, we arbitrarily defined three subsets of characters for these keys including (1) only characters of very straightforward use (included in the subset "step1: use only the most straightforward characters to get a short list of candidate species"), (2) all characters except the ones of most difficult use

(subset "step 2: try identification by excluding only the most difficult characters") and (3) all characters, including "easy", "average", and "difficult" ones (subset "step 3: use also difficult characters if step 2 does not bring to species identification"). The user has the possibility of following a three steps identification procedure that considers characters of straightforward use at first, followed by characters of more and more difficult interpretation. This procedure should facilitate identification and reduce the risk of misidentification (particularly if a species can be identified only through step 1 or through step 1 and 2). We also defined a subset for species of economical importance. The use of this subset should speed up the identification of the more commonly trapped / intercepted taxa. When using this subset, identification should be carefully verified *a posteriori* (through the hyperlink to species description, see below) as all the less common species not included in this subset might be erroneously identified as species of economical importance (false positives). Of course, character and species subsets can all be ignored and the user can either arbitrarily score any of the characters available from the full list or use the "best" option provided by the LUCID software which should allow choosing characters with the highest discrimination power (the "best" option can be repeatedly used after eliminating redundant characters through the "prune" option). In any case, being a multi-entry key, the user can always decide either to skip characters, to choose multiple answers whenever he is uncertain about the correct score and/or to restrict the identification only to the most common species.

We tried to make the technical terminology used in the single-entry keys more accessible to non-specialists by adopting a consistent framework of character names and indicating in parentheses alternative names of the same character in the published scientific literature (as it happens for example with the *Ceratitis* subapical / cubital / preapical wing band). We then embedded images that clearly illustrate name and position of each character on the insect body as well as images showing how the same character state looks in different species. An initial set of 2300 images was assembled from the databases of the Royal Museum for Central Africa (RMCA) and of the London Natural History Museum (NHM). Images were grouped according to species name and body part (head, thorax dorsal, thorax lateral, abdomen, wings, legs), divided in groups and, when possible, assigned to each combination of character state and species name. This generated a database of approximately 28000 repeated images (for example, the same thorax image of a particular species was repeatedly used to illustrate postpronotal lobe, scutum and scutellum characters for that species). The large set of embedded images aims at clearly illustrating the morphological variability of the same character state across species. In fact, we consider that many terms used to describe morphological variation (such as "small / large, darker / paler, thicker / thinner etc.") while being straightforward for a tephritid taxonomist (who can rely on the experience accumulated after the examination of large numbers of specimens) are not always clear to non-specialised users. Therefore, we dedicated particular attention to provide multiple images to show, for example, how "narrow" a wing discal band should be, before being considered as "broad" or how "small" a postpronotal spot can be before being scored as "occupying most of postpronotum".

Once a tentative identification is obtained (or when the list of candidate species is reduced to a few taxa), the keys give the possibility of verifying the correspondence between the examined voucher and (1) the species description as it appears in the published scientific literature and (2) images from the RMCA and NHM tephritid collections. Discrepancies between the examined voucher and available images (as it might result from the occurrence of multiple character states for a species) can then be verified through hyperlinks to either the species description or to all character states considered for that species in the LUCID input matrix. Information regarding the taxonomic status, geographic distribution and collection specimens of each taxon is also available through hyperlinks to Encyclopedia of Life (EOL) and to the Belgian Biodiversity Platform (BeBIF, a section of GBIF, the Global Biodiversity Information Facility). Links to the Barcoding of Life Database website (BOLD) allow verifying the availability and geographical coverage of DNA barcodes for each species. In some cases, the available character list will not always allow the unambiguous identification of a taxon (as it happens, for example, with females of the subgenus *Ceratitis* (*Pterandrus*)). Under these circumstances, the direct comparison of species descriptions and distributions is the best strategy to try and resolve the short list of candidate taxa.

The keys can be accessed online (http://keys.lucidcentral.org/keys/v3/fruitflies/) or freely downloaded and used from a computer hard drive (supplementary files SF3-10). The first option is only recommended for a preliminary overview of the key structure, while downloading and running the keys (e.g. from a memory stick used as a removable device) should allow a faster and more effective use of the software. A quick start guide providing basic information about the key functioning is associated to the downloadable version.

Acknowledgments

This work has been co-funded by the Belgian Directorate-General for Development Cooperation (through framework agreement with the Royal Museum for Central Africa) and by the International Atomic Energy Agency (IAEA - Vienna, project "Development of a Web Based Multi Entry Key for Fruit Infesting Tephritidae", contract n. 16859). The last author greatly acknowledges travel grants of the Research Foundation - Flanders (FWO-Vlaanderen) for study visits to the Natural History Museum (London, UK), the Plant Protection Research Institute (Pretoria, South Africa), and the International Institute of Tropical Agriculture (Cotonou, Benin) to examine specimens in preparation of the character matrices. An earlier version of the *Ceratitis* and *Trirhithrum* keys were developed through the U.S. Agency for International Development (USAID, PCE-G-00-98-0048-00) and the U.S. Department of Agriculture (USDA) / the National Institute of Food and Agriculture (CSREES) / the Initiative for Future Agricultural and Food Systems (IFAFS) grants to Texas A&M University (00-52103-9651). We are grateful to Alain Reygel (RMCA - Tervuren) and to Georg Goergen (International Institute of Tropical Agriulture - Cotonou) for their contribution to the image dataset as well as to Myriam Vandenbosch for practical and moral support.

References

Aluja M, Mangan RL (2008) Fruit fly (Diptera: Tephritidae) host status determination: critical conceptual, methodological, and regulatory considerations. Annual Review of Entomology 53: 473–502. doi: 10.1146/annurev.ento.53.103106.093350

Aluja M, Norrbom A (1999) Fruit flies (Tephritidae) phylogeny and evolution of behavior. CRC Press, Boca Raton, 967 pp. doi: 10.1201/9781420074468

Armstrong KF, Ball SL (2005) DNA barcodes for biosecurity: invasive species identification. Philosophical Transactions of the Royal Society B: Biological Sciences 360: 1813–1823. doi: 10.1098/rstb.2005.1713

Carvalho M, Bockmann Fv, Amorim D, Brandão C, Vivo Mr, Figueiredo JL, Britski H, Pinna MrC, Menezes Nr, Marques FL, Papavero N, Cancello E, Crisci J, McEachran J, Schelly R, Lundberg J, Gill A, Britz R, Wheeler Q, Stiassny MJ, Parenti L, Page L, Wheeler W, Faivovich Jn, Vari R, Grande L, Humphries C, DeSalle R, Ebach M, Nelson G (2007) Taxonomic impediment or impediment to taxonomy? A commentary on systematics and the cybertaxonomic-automation paradigm. Evolutionary Biology 34: 140–143. doi: 10.1007/s11692-007-9011-6

De Meyer M (1996) Revision of the subgenus *Ceratitis* (*Pardalaspis*) Bezzi, 1918 (Diptera, Tephritidae, Ceratitini). Systematic Entomology 21: 15–26. doi: 10.1111/j.1365-3113.1996.tb00596.x

De Meyer M (1998) Revision of the subgenus *Ceratitis* (*Ceratalaspis*) Hancock (Diptera: Tephritidae). Bulletin of Entomological Research 88: 257–290. doi: 10.1017/S0007485300025888

De Meyer M (2000) Systematic revision of the subgenus *Ceratitis* Macleay *s.s.* (Diptera, Tephritidae). Zoological Journal of the Linnean Society 128: 439-467. doi: 10.1111/j.1096-3642.2000.tb01523.x

De Meyer M (2006) Systematic revision of the fruit fly genus *Carpophthoromyia* Austen (Diptera, Tephritidae). Zootaxa 1235: 1–48.

De Meyer M (2009) Taxonomic revision of the fruit fly genus *Perilampsis* Bezzi (Diptera, Tephritidae). Journal of Natural History 43: 2425–2463. doi: 10.1080/00222930903207868

De Meyer M, Freidberg A (2005) Revision of the fruit fly genus *Capparimyia* (Diptera, Tephritidae). Zoologica Scripta 34: 279–303. doi: 10.1111/j.1463-6409.2005.00195.x

De Meyer M, Freidberg A (2006) Revision of the Subgenus *Ceratitis* (*Pterandrus*) Bezzi (Diptera: Tephritidae). Israel Journal of Entomology 36: 197–315. http://www.entomology.org.il/sites/default/files/pdfs/p.197 [de meyer and freidberg.pdf]

De Meyer M, Freidberg A (2012) Taxonomic revision of the fruit fly genus *Neoceratitis* Hendel (Diptera: Tephritidae). Zootaxa 3223: 24–39.

De Meyer M, Robertson MP, Peterson AT, Mansell MW (2008) Ecological niches and potential geographical distributions of Mediterranean fruit fly (*Ceratitis capitata*) and Natal fruit fly (*Ceratitis rosa*). Journal of Biogeography 35: 270–281. doi: 10.1111/j.1365-2699.2007.01769.x

Malacrida A, Gomulski L, Bonizzoni M, Bertin S, Gasperi G, Guglielmino C (2007) Globalization and fruitfly invasion and expansion: the medfly paradigm. Genetica 131: 1–9. doi: 10.1007/s10709-006-9117-2

Norrbom AL (2004) Updates to biosystematic database of world diptera for tephritidae through 1999. Diptera Data Dissemination Disk. [CD]

Smit J, Reijnen B, Stokvis F (2013) Half of the European fruit fly species barcoded (Diptera, Tephritidae); a feasibility test for molecular identification. ZooKeys 365: 279–305. doi: 10.3897/zookeys.365.5819

Virgilio M, Backeljau T, Nevado B, De Meyer M (2010) Comparative performances of DNA barcoding across insect orders. BMC Bioinformatics 11: 1–10. doi: 10.1186/1471-2105-11-206

Virgilio M, Jordaens K, Breman FC, Backeljau T, De Meyer M (2012) Identifying insects with incomplete DNA barcode libraries, African fruit flies (Diptera: Tephritidae) as a test case. PLoS ONE 7: e31581. doi: 10.1371/journal.pone.0031581

White I, Copeland R, Hancock D (2003) Revision of the afrotropical genus *Trirhithrum* (Diptera: Tephritidae). Cimbebasia 18: 71–137.

White IM (2006) Taxonomy of the Dacina (Diptera:Tephritidae) of Africa and the Middle East. African Entomology Memoir 2: 1–156.

White IM, Goodger KFM (2009) African *Dacus* (Diptera: Tephritidae); new species and data, with particular reference to the Tel Aviv University collection. Zootaxa 2127: 1–49.

Wilson MR (2000) Loss of taxonomists is a threat to pest control. Nature 407: 559–559. doi: 10.1038/35036751

Supplementary material 1

List of species, characters and character states considered in each identification key

Authors: Massimiliano Virgilio, Ian White, Marc de Meyer

Data type: multimedia

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: doi: 10.3897/zookeys.428.7366.app1

Supplementary material 2

List of positive and negative character dependencies in each identification key

Authors: Massimiliano Virgilio, Ian White, Marc de Meyer

Data type: multimedia

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: doi: 10.3897/zookeys.428.7366.app2

Supplementary material 3

Key to genera

Authors: Massimiliano Virgilio, Ian White, Marc de Meyer

Data type: multimedia

Explanation note: A set of multi-entry identification keys to African frugivorous flies (Diptera, Tephritidae): key to genera.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: doi: 10.3897/zookeys.428.7366.app3

Supplementary material 4

Key to *Capparimyia*
Authors: Massimiliano Virgilio, Ian White, Marc de Meyer
Data type: multimedia
Explanation note: A set of multi-entry identification keys to African frugivorous flies (Diptera, Tephritidae): key to *Capparimyia*.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: doi: 10.3897/zookeys.428.7366.app4

Supplementary material 5

Key to *Carpophthoromyia*
Authors: Massimiliano Virgilio, Ian White, Marc de Meyer
Data type: multimedia
Explanation note: A set of multi-entry identification keys to African frugivorous flies (Diptera, Tephritidae): key to *Carpophthoromyia*.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: doi: 10.3897/zookeys.428.7366.app5

Supplementary material 6

Key to *Ceratitis*

Authors: Massimiliano Virgilio, Ian White, Marc de Meyer

Data type: multimedia

Explanation note: A set of multi-entry identification keys to African frugivorous flies (Diptera, Tephritidae): key to *Ceratitis*.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: doi: 10.3897/zookeys.428.7366.app6

Supplementary material 7

Key to *Dacus* and *Bactrocera*

Authors: Massimiliano Virgilio, Ian White, Marc de Meyer

Data type: multimedia

Explanation note: A set of multi-entry identification keys to African frugivorous flies (Diptera, Tephritidae): key to *Dacus* and *Bactrocera*.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: doi: 10.3897/zookeys.428.7366.app7

Supplementary material 8

Key to *Neoceratitis*

Authors: Massimiliano Virgilio, Ian White, Marc de Meyer

Data type: multimedia

Explanation note: A set of multi-entry identification keys to African frugivorous flies (Diptera, Tephritidae): key to *Dacus* and *Neoceratitis*.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: doi: 10.3897/zookeys.428.7366.app8

Supplementary material 9

Key to *Perilampsis*

Authors: Massimiliano Virgilio, Ian White, Marc de Meyer

Data type: multimedia

Explanation note: A set of multi-entry identification keys to African frugivorous flies (Diptera, Tephritidae): key to *Dacus* and *Perilampsis*.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: doi: 10.3897/zookeys.428.7366.app9

Supplementary material 10

Key to *Trirhithrum*

Authors: Massimiliano Virgilio, Ian White, Marc de Meyer

Data type: multimedia

Explanation note: A set of multi-entry identification keys to African frugivorous flies (Diptera, Tephritidae): key to *Dacus* and *Trirhithrum*

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: doi: 10.3897/zookeys.428.7366.app10